Multimodal single-cell sequencing of the human cortex reveals neuronal vulnerability and activated glial cell states in focal cortical dysplasia

Isabella C. Galvão[1,2], Manuela Lemoine[1,2], Lauana A. Messias[1,2], Patrícia A.O.R.A. Araújo[1,2], Jaqueline C. Geraldis[1,2], Clarissa L. Yasuda[2,3], Marina K. M. Alvim[2,3], Enrico Ghizoni[2,3], Helder Tedeschi[2,3], Fernando Cendes[2,3], Fabio Rogerio[2,4], Iscia Lopes-Cendes[1,2], and Diogo F. T. Veiga[1,2,*]

[1]Department of Translational Medicine, School of Medical Sciences, University of Campinas, Campinas (UNICAMP), Brazil

[2]The Brazilian Institute of Neuroscience and Neurotechnology (BRAINN), Campinas, Brazil

[3]Department of Neurology, School of Medical Sciences, University of Campinas (UNICAMP), Campinas, SP, Brazil

[4]Department of Pathology, School of Medical Sciences, University of Campinas (UNICAMP), Campinas, Brazil

[*]Corresponding author: dfveiga@unicamp.br

ABSTRACT

The role of the diverse brain cell types in focal cortical dysplasia (FCD) development and epileptogenicity is not fully understood. Here, we performed multi-omics single-cell sequencing to characterize the chromatin accessibility and transcriptomes of cortical lesions and non-lesion areas from individuals with FCD IIa and IIb, the most prevalent presentations of this neurodevelopmental disorder. Our integrative analyses of the multimodal cell atlas, encompassing 61,525 nuclei with paired open chromatin and gene expression, uncovered a selective loss of upper-layer excitatory neurons and identified a lesion-specific population expressing NEFM, a neurofilament associated with dysmorphic neurons. In glial compartments, we observed a shift towards immature astrocytic populations exhibiting features of reactive astrocytes and balloon cells. Moreover, we identified activated microglial cell states emerging in FCD IIb, indicating that neuroinflammation contributes to FCD pathogenesis. This multimodal cell atlas is a valuable resource for exploring the cellular landscape and informing novel therapeutic approaches for cortical malformations.

Keywords: focal cortical dysplasia; cortical malformations; single-cell sequencing; multi-omics; gene expression; chromatin accessibility; microglia; astrocyte differentiation

INTRODUCTION

45 Focal Cortical Dysplasia (FCD) is a malformation of the cerebral cortex and a major cause of drug-refractory epilepsy in children. The treatment of this neurological condition with current anti-seizure medications is often ineffective, potentially demanding surgical removal of 48 affected brain tissue. FCD is classified based on neuroimaging, histopathology, and genetic characteristics[1,2]. FCD type II, the most common presentation, is characterized by cytoarchitectural cortical abnormalities including loss of cortical lamination, blurring of the 51 gray-white matter junction, and abnormal cellular development[2]. FCD II cortical lesions can be further classified based on cytological features: FCD IIa is characterized by the presence of dysmorphic neurons, while FCD IIb presents with both dysmorphic neurons and balloon cells. 54 Dysmorphic neurons are abnormal cells featuring enlarged soma and nuclei[1,2], along with expression of neurofilament markers typical of immature neurons in the developing cortex. On the other hand, balloon cells are characterized by large cell bodies with vitreous, opalescent, 57 and eosinophilic cytoplasm[1,2]. These cells express markers commonly associated with neuroglial progenitors, including nestin and vimentin, indicating their immature nature[3]. High expression of other progenitor cell markers such as c-Myc, SOX2, and Oct-4 were also 60 observed throughout FCD II lesions[4,5].

FCD II is also characterized by activation of the mammalian target of rapamycin complex 1 (mTORC1) signaling pathway, which is implicated in cell growth, increased 63 synaptic transmission and plasticity, as well as regulation of immune system responses[6–8]. Hyperactivation of mTOR signaling, which operates through the mTORC1 and mTORC2 complexes, is particularly observed in abnormal cells such as dysmorphic neurons and balloon 66 cells[9]. Increased phosphorylation of S6K1 and S6 proteins, which are mTORC1 substrates, can be used as a biomarker to distinguish FCD type II lesions[4]. Genetic alterations such as somatic

gain-of-function mutations in mTOR activating genes and germline loss-of-function mutations of mTOR inhibitors are also found in FCD II[9–11].

The cellular and epileptogenic mechanisms of FCD are not yet fully understood. Dysmorphic neurons appear to play a pivotal role in generating epileptic discharges[11], while balloon cells are considered electrically quiescent and may even have an anti-seizure effect[3]. A recent study showed that dysmorphic neurons derive from glutamatergic neurons while balloon cells have astrocytic lineage[12]. We and other groups have shown that FCD IIb lesions are also characterized by increased microglia and astrocyte reactivity when compared to FCD IIa or healthy controls[13–15]. However, little is known regarding the cellular landscape of FCD lesions and the role of the diverse cortical cell types in the disease.

Single-cell assays enable the capturing of individual cells or nuclei for molecular profiling of diverse biological modalities, including gene expression and open chromatin levels. Single-nucleus RNA sequencing enables measuring gene expression, while transposase-accessible chromatin (ATAC) sequencing measures chromatin accessibility, which indicates the physical access to DNA for transcription. These assays have been applied to resolve the cellular diversity of the human brain in healthy and disease contexts[16–18] and to discover novel cell states associated with neurological conditions[19–21]. In addition, recently introduced multimodal single-cell approaches can obtain multiple measurements from the same cells simultaneously, and integration of these layers might yield even richer information regarding the cellular mechanisms involved in diseases[22,23].

In this study, we applied multi-omics single-cell sequencing to characterize the chromatin accessibility and gene expression profiles in FCD type II, generating a comprehensive multimodal atlas from individuals diagnosed with the condition. This resource was explored to unravel molecular processes and cell states underlying FCD pathogenesis.

93 **RESULTS**

Assembling a multimodal cellular atlas of FCD type II

We performed multi-omics single-cell sequencing of cortical brain tissue from FCD IIa (n=3)
96 and IIb (n=6) lesions, as well as from non-lesion tissues from FCD donors selected as controls (n=2; Figure 1A). Specimens were collected during epilepsy surgery to treat drug-refractory FCD. Internal controls correspond to histologically normal cortical tissues where neither
99 abnormal cells (dysmorphic neurons and balloon cells) nor architectural changes were observed. Dissociated nuclei were profiled using the 10X Genomics Multiome ATAC + Gene Expression assay for simultaneous measurement of chromatin accessibility and gene
102 expression in individual nuclei. Quality control was performed to remove low-quality nuclei and to filter out potential doublets, and Harmony[24] was applied to ATAC and RNA datasets to integrate samples from different sequencing batches (Figure 1B). The Weighted Nearest
105 Neighbor (WNN) approach[25] from Seurat was used to create an integrated multimodal (ATAC + RNA) visualization (Figure 1B). The resulting multimodal FCD atlas comprised 61,525 nuclei with paired chromatin and gene expression measurements. Clustering analysis in the
108 multimodal dataset identified a total of 40 nuclei clusters (Figure S1A), which were annotated using Azimuth[25] followed by manual validation using canonical marker genes to define consensus cell types. Azimuth was able to assign labels with high confidence to the majority
111 of clusters (Figure S1B), and none of the clusters were exclusive to a unique sample (Figure S1C). Consensus cell type annotation was performed using a broad classification (Figure 1C) and higher-resolution taxonomy including neuronal subtypes and cortical layers (Figure 1D).

114 Overall, our annotation identified the various cell types expected in the human cortex such as excitatory and inhibitory neurons, astrocytes, oligodendrocytes, oligodendrocytes progenitor cells (OPC), microglia, endothelial cells, and vascular and leptomeningeal cells
117 (VLMC, Figure 1C). We also identified the various subtypes of excitatory neurons (EN, Figure

1D) including upper layers EN L2/3 (termed EN Upper), and deep layer subtypes EN L5, EN L5/6, EN L6 (EN Deep), while inhibitory neurons (IN) were further classified into Vip, Sst, Pvalb, Lamp5 and Sncg/Lamp5 subtypes (Figure 1D). Based on the histological assessment, most of the sequenced nuclei derived from IIb lesions, followed by internal controls and IIa lesions (Figure 1E). Gene expression and open chromatin accessibility of marker genes were inspected to confirm cell type identity (Figures 1F-G). For instance, *AQP4* (astrocytes), *SLC17A7* (EN), *GAD2* (IN), *PLP1* (oligodendrocytes), *CD86* (microglia), and *GPR17* (OPC) were marker genes with concordant expression and chromatin accessibility in their corresponding cell types.

Using signature enrichment analysis, mTORC1 signaling was found to be elevated in dysplastic tissue across all cell types except VLMC and endothelia (Figure S2A), and cells with the highest mTORC1 activation were identified in FCD IIb (Figure S2B). Thus, the single-cell multimodal atlas uncovered the heterogenous composition of FCD type II lesions and recapitulated key aspects of their pathological state.

Unraveling cellular changes in FCD lesions

Next, we determined cell type abundance changes between FCD and control samples with Speckle[26], using an approach that accounts for variability in cellular frequencies among patients. As a result, we found several alterations between lesions and controls, including neuronal loss in FCD affecting the EN L2/3 subtype, and increased frequency of oligodendrocytes in both FCD IIa and IIb (Figure 1H). In addition, we found an expansion of the microglia population specifically in IIb donors.

We also leveraged the hierarchical nature of the Azimuth annotation, which classifies glutamatergic and GABAergic neurons into 116 subclasses based on the expression of makers, to find out how these neuronal subclasses are affected in the disease. The Exc L2

LINC00507/*GLRA3*, the most common EN subclass in the cortex in our samples, was also the
144 most affected glutamatergic neuron displaying the highest number of transcriptional changes and abundance reduction in lesions (Figure S3A,B). Among IN, FCD IIa displayed an increased frequency of Inh L1-5 Vip/*SMOC1*, a subclass of vasoactive intestinal peptide-
147 expressing (Vip) interneurons, and the Parvalbumin-expressing (Pvalb) interneurons Inh L3 Pvalb/*SAMD13* (Figure S3C); yet, these were not the subclasses with most transcriptional changes (Figure S3D). On the other hand, somatostatin-expressing (Sst) subclasses Inh L1
150 Sst/*DEFB108B* and Inh L1-2 Sst/*CCNJL* were decreased only in FCD IIb when compared to controls (Figure S3C). Altogether, this data shows that cellular abundance changes affect both neuronal and non-neuronal populations in FCD pathogenesis.

153

Identification of a disease-specific cluster containing dysmorphic neurons

During annotation, we found a heterogeneous cluster containing both excitatory neurons and
156 oligodendrocytes that we initially termed Unclassified since Azimuth could not confidently assign a cell type (Figure 1C). This cluster was highly disease-specific as it contained 1,308 cells, being that 1,285 cells (98%) originating from FCD (Figure 2A). Despite clustering next
159 to oligodendrocytes after Harmony integration, 54% of nuclei were classified as EN and 29% as oligodendrocytes. To further characterize this cell population, we identified marker genes using Seurat´s FindMarkers and performed KEGG pathway enrichment analysis on the set of
162 marker genes using g:Profiler[27]. Notably, all top enriched pathways were related to neurodegenerative diseases including Parkinson's, amyotrophic lateral sclerosis, and Alzheimer's (Figure 2B). Also, the majority of DS cluster markers were associated with
165 neuronal processes including oxidative phosphorylation, Ca^{2+} transport, neuron development, and neurofilament subunits (Figure 2C).

We found two neurofilament subunits *NEFL* and *NEFM* overexpressed in the DS
168 cluster (Figure 2C), and that *NEFM* is exclusively expressed in the DS cluster and not detected in any other cell type in either lesions or non-lesion tissues (Figure 2D). The chromatin accessibility in the *NEFM* and *NEFL* loci corroborates with their active gene expression in the
171 DS cluster (Figure 2E). In particular, SMI32 or SMI-311 antibodies against neurofilaments are often used to identify dysmorphic neurons in dysplastic tissue during histopathological diagnosis of FCD[2,11].

174 Next, we used a signature enrichment analysis to further characterize the DS cluster. We used a 30-gene signature from Baldassari et al[28], which obtained a purified DN population by isolating these cells in FCD lesions using laser capture microdissection followed by single-
177 cell sequencing. We applied the AddModuleScore function from Seurat to score the DN signature and found that nuclei with the highest expression of this signature were overlapping the DS cluster (Figure 1F). Finally, to find out whether the DS cluster could be detected in
180 healthy individuals, we integrated our FCD dataset with single-cell RNA-seq from cortical samples obtained from age-matched autopsy controls[17] using Harmony and carried out cell annotation. As a result, we confirmed that the DS cluster was not present in the normal cortex
183 (Figure 1G). Taken together, these data demonstrate that this cell population is uniquely present in FCD lesions and comprises dysmorphic neurons expressing genes associated with neurodegenerative conditions.

186

Pathological microglia cell states in FCD

To further characterize microglia states in FCD, we performed subclustering and re-annotation
189 of the broad microglial compartment. Subclustering and marker detection with Seurat uncovered 7 subpopulations with clearly distinct gene signatures (Figure S4; full list of subcluster markers provided in Supplemental Table 2). Diverse microglia signatures identified

192 by single-cell sequencing have been previously reported in neurological diseases[20,21,29]. We used signatures of microglia subsets derived from healthy and epileptic temporal cortex[21] to annotate subclusters in FCD. Using this approach, we identified three distinct microglia states

195 in FCD lesions (Figure 3A): a homeostatic subset (the largest population corresponding to subcluster 0); a subset expressing high levels of *CD74* (subcluster 1, *CD74*$^{+,}$), and a subset expressing high levels of *CD83* (subclusters 2 and 5, *CD83*$^{+}$). We also identified small

198 subclusters of lymphocyte/T cells, B cells, and perivascular macrophages that were annotated using classical lymphocyte markers (Figure 3A). As shown in Figure 3B, *CD74*$^{+}$ microglia expressed marker genes related to antigen processing and presentation/MHC class II genes

201 (*HLA-DRA*, *HLA-DPA1*, *HLA-DPB1*, *HLA-DRB1*, *HLA-DQA1*), adaptive immune system response and microglial cell activation (*CD74*, *SC1N*, *C1QA*, *C1QB*, *C1QC*), and interferon signaling (*IFNGR1*, *JAK2*, *B2M*). On the other hand, *CD83*$^{+}$ microglia expressed high levels

204 of pro-inflammatory cytokines including *IL1B*, *CCL2*, *CCL3*, *CCL4*, and *TNFSF18*. Interestingly, *CD83* and *CD74* expression are largely non-overlapping, indicating that these surface markers can be effectively used to distinguish microglia states in FCD (Figure 3C).

207 Of note, homeostatic microglia and activated states are expanded only in FCD IIb tissues (Figure 3D). Compared to controls, FCD IIb had a higher frequency of *CD74*$^{+}$ and homeostatic subclusters. Compared to FCD IIa, the expansion included *CD74*$^{+}$ and *CD83*$^{+}$

210 subpopulations as well as lymphoid/T cells (Figure 3D). Together, these data highlight the emergence of activated microglial states specifically within the FCD IIb lesion environment.

213 **Open chromatin and gene expression microglia activation in FCD type IIb**

Next, we found differentially expressed genes in FCD IIb microglia by performing a Wilcox test after SCT transformation[30] with correction for sequencing batch, mitochondrial gene

216 percentage, and brain region. The analysis identified 133 upregulated genes in FCD IIb

compared to controls (Figure 3E). GO terms associated with upregulated genes included antigen processing and presentation, MHC class II protein complex, and positive regulation of
219 cell activation (Figure 3F), in line with the expansion of *CD74*$^{+}$ microglia in FCD IIb.

We then applied chromVAR[31] to measure motif activity based on the open chromatin profiles in individual nuclei and determined the set of motifs with differential activity between
222 FCD IIb and controls. Top upregulated motifs in FCD IIb microglia included ELF1, IRF8, and STAT1/STAT2 (Figure 3G), which are regulators related to microglia activation. Cells with the highest motif activity in FCD IIb overlapped the *CD74*$^{+}$ microglia (Figure 3H), denoting
225 that these microglia undergo concerted transcriptional and epigenetic changes in response to the disease. Therefore, gene expression and motif activity in open chromatin regions indicate the microglia activation in FCD IIb is mediated by the *CD74*$^{+}$ subpopulation.

228

Immature astrocyte cell states and balloon cells in FCD

We explored whether the astrocytic population is affected in FCD by employing an
231 unsupervised trajectory analysis. Astrocytes were clustered and temporally ordered on a trajectory as a function of pseudotime using Monocle3, and gene expression changes were visualized over UMAPs (Figure 4A). As a result, the trajectory built by Monocle3 organized
234 astrocytes according to their differentiation status, with immature and mature astrocytes on opposite ends of the trajectory (Figure 4B). The root of the trajectory was chosen based on the expression of immature (*CD44*, *TNC*, *VCAN*, and *GFAP*) and mature (*SLC1A2*, *SLC1A3*, and
237 *GPC5*) astrocyte markers. Remarkably, the quantification of FCD and control cells along the trajectory revealed an expansion of immature astrocytes accompanied by a decrease of mature astrocytes in FCD tissues (Figure 4C). On the other hand, fully mature astrocytes were highly
240 represented in controls (Figure 4C). We also analyzed the expression of gene signatures corresponding to balloon cells, which have been shown to originate from the astrocytic

lineage[28] as well as reactive astrocytes[19], using the AUCell method[32] and the AddModuleScore
243 function from Seurat. Scores for both balloon cells and reactive astrocytes signatures were more elevated in astrocytes from FCD lesions and overlapped with immature states at the beginning of the trajectory (Figure 4D). Of note, immature astrocytes identified as balloon cells
246 were mostly exclusive to FCD IIb cases, while reactive astrocytes were present in both FCD subtypes (Figure 4D, bottom). Genes associated with balloon cells and reactive astrocytes including *CHI3L1*, *HSPB1*, *SERPINA3*, *GFAP*, and *VIM* were up-regulated in FCD IIb
249 astrocytes compared to FCD IIa or internal controls (Figure 4E). Taken together, these analyses showed that immature astrocyte cell states, including balloon cells and reactive astrocytes, are expanded in the disease, indicating that normal astrocyte differentiation is impaired in FCD.

252

Chromatin changes and cell-type specific gene regulation in FCD

Next, we explored the multimodal atlas to find open chromatin regions associated with gene
255 regulation in FCD. First, we found the set of differentially accessible chromatin regions (DACRs), which are regions (peaks) with gain or loss of chromatin accessibility, by comparing chromatin profiles across cell types in FCD against controls. Overall, FCD IIa had a higher
258 number of DACRs, especially in upper and deep EN (Figure 5A). Most of the DACRs in FCD IIa displayed a gain of accessibility in the lesion (opening peaks), while DACRs in FCD IIb were characterized by a loss of chromatin accessibility (Figure 5A). Further analysis using
261 Encode annotation found that most DACRs were located within enhancer regions, with 34% overlapping proximal enhancers near a transcriptional start site and 44% overlapping distal enhancers (Figure 5B).

264 We then sought to understand whether these DACRs were driving gene regulation. To achieve this, we implemented a machine-learning approach based on random forests to learn regulatory links using the paired single-cell chromatin and gene expression readouts in the

267 dataset (see Methods). The method was applied to find chromatin regions that were most likely to predict gene expression in each cell type, thus revealing the role of DACRs in controlling gene expression. For instance, we identified regulatory links for *FKBP* prolyl isomerase 5
270 (*FKBP5*), a gene within the mTORC1 pathway that exhibited upregulation in FCD astrocytes (Figure 5C). The regulatory connections involved DACRs with gain of chromatin accessibility in FCD (regions highlighted in Figure 5C), suggesting that the increased *FKBP5* expression is
273 accompanied by chromatin remodeling. These opening chromatin peaks are also putative *FKBP5* cis-regulatory elements since they are located in regions with enhancer activity according to Encode. Similarly, we also found opening chromatin peaks located in enhancers
276 that were linked to *C1QA* and *C1QB*, which were genes involved in microglia cell activation increased in FCD IIb (Figure S5). Taken together, these analyses demonstrated that chromatin changes affecting enhancer regions are likely driving cell type-specific gene regulation in FCD.
279

282

DISCUSSION

285 In this work, we constructed a multimodal cellular atlas of FCD, a cortical malformation prevalent in children and adults and a leading cause of drug-resistant focal epilepsy. This multi-omics dataset contains paired chromatin accessibility and gene expression of FCD II lesions
288 and adjacent normal tissue, and will be a valuable resource for exploring the molecular and cellular mechanisms of FCD pathogenesis. Thus, we anticipate that this multimodal cell atlas will be important in future studies integrating chromatin and transcriptomic data, including
291 unraveling the impact of non-coding variants in open chromatin regions in gene expression or identifying chromatin enhancers driving gene dysregulation.

The majority of single-cell studies so far have focused on a single modality such as
294 RNA or ATAC sequencing, and a few studies have analyzed the human brain from a multi-omics perspective[16,33]. Paired single-cell studies, which generate multiple readouts from the same cell, are still in their infancy due to technical limitations and cost. Yet, these paired
297 multimodal single-cell experiments generate unique datasets that enable exploration of cell-type specific gene regulation by learning the relationship between chromatin accessibility and transcription in the same cell[34].

300 Herein, we leveraged the multimodal FCD atlas to reveal the concerted epigenetic and transcriptional alterations affecting both neuronal and glial populations in response to the disease. Using complementary single-cell methods to explore the FCD multimodal cell atlas in
303 detail, we discovered pathological cell states associated with FCD development and characterized the involvement of novel glial and neuronal subpopulations in response to FCD.

First, our data showed that both glutamatergic and GABAergic neurons undergo
306 changes in cellular abundance and transcriptional profiles during FCD development, yet neuronal loss affects mostly the EN L2/3 subtype. Interestingly, EN in upper cortical layers

were also found to be reduced in cortical samples of patients with multiple sclerosis[35] and
309 temporal lobe epilepsy[36], highlighting their vulnerability in other neurological conditions.

Second, we characterized a neuronal cell population specific to FCD lesions expressing a dysmorphic neuron-signature and neurofilament genes. This neuronal population was notably
312 absent in non-lesion controls and the healthy cortex from autopsy cases. Also, the expression of *NEFM*, encoding the medium polypeptide subunit of the neurofilament protein, was found to be restricted to this subpopulation. Neurofilaments detected in the blood have emerged as
315 reliable diagnostic biomarkers in several neurological diseases[37]. The remarkable specificity of *NEFM* expression in dysplastic cells suggests its potential as a blood biomarker for FCD, prompting the need for additional research to explore its utility as a molecular diagnostic tool
318 in the disease.

Third, we described the expansion of microglia and the emergence of pathological cell states specifically in FCD IIb lesions. We identified two FCD-associated microglial
321 subpopulations with distinct signatures: a *CD74*$^{+}$ subset overexpressing antigen presentation molecules (MHC class-II components), and a *CD83*$^{+}$ subset with increased expression of pro-inflammatory cytokines. Previous scRNA-seq studies have identified microglial populations in
324 neurological conditions[20,21,29]. For instance, using scRNA-seq followed by in situ immunohistochemistry, Olah et al[21] described the existence of nine microglial subpopulations in the human cortex (prefrontal cortex of autopsy samples and temporal cortex of epilepsy),
327 including the *CD83*$^{+}$ and *CD74*$^{+}$ subsets. Mathys et al[38] also identified a microglia subset with increased expression of *CD74* and MHCII-related genes in patients with Alzheimer's disease. A *CD74* subset was also described in the pathology of multiple sclerosis[39], and *CD74* has been
330 proposed as a marker for reactive microglia due to its increased expression levels in disease[40]. Although there is no consensus in the field on how to annotate microglia cell states, we used the terminology from Olah et al since the *CD83* and *CD74* molecules were identified

333 as markers of microglia subclusters in FCD. Here, we also found higher lymphoid/T-cells infiltration is specific to FCD IIb, in agreement with previous findings[14].

There is also strong evidence of dysregulated neuroinflammation mediated by 336 microglia in FCD and epilepsy. Zimmer et al[14] showed that FCD IIb elicits a stronger inflammatory response characterized by increased expression of the *HLA-II* and *CCL2* in microglia near balloon cells. Kumar et al[41] described immune cell infiltration and increased 339 expression of pro-inflammatory genes (e.g. *IL1B*, *CCL4*, *CCL2*) in microglia from epilepsy samples. Therefore, we postulate that the $CD74^{+}$ and $CD83^{+}$ microglia subsets identified herein are the glial populations mediating immune system activation and neuroinflammation in FCD 342 IIb.

Finally, we showed that astrocyte differentiation is impaired in FCD, and immature astrocytic cell states are expanded in lesions. This shift from mature to immature astrocytes 345 can be explained in part by the appearance of reactive astrocytes and balloon cells, which are abnormal cell populations typically found in the lesion microenvironment, that we characterized as immature astrocytic states. Specifically, pseudotime trajectory analysis 348 revealed that reactive astrocytes resemble immature-like cells that emerge in FCD lesions, supporting the view that astrogliosis induces a partial reversion of astrocyte maturation[42]. Similarly, we showed that balloon cells also represent immature astrocytic cell states. Balloon 351 cells were identified exclusively in FCD IIb, consistent with the histological characterization of the clinical samples. Moving forward, it will be important to understand the contribution of these immature astrocytes to FCD. Mature astrocytes are known to maintain neuronal 354 homeostasis and prevent excessive neuronal activity through neurotransmitter reuptake[42], and their dysregulation can lead to epileptogenesis[43]. Thus, these immature astrocytic populations might add to the intrinsic epileptogenicity in FCD.

357

METHODS

Clinical samples and neuropathological diagnosis

360 Fresh brain samples were collected from patients who underwent surgery for intractable seizures after clinical diagnosis of FCD at Hospital de Clínicas, University of Campinas. Surgical specimens were (i) formalin-fixed paraffin-embedded (FFPE) or (ii) snap-frozen in 363 liquid nitrogen and stored at −80°C. FFPE samples were submitted to diagnostic routine in serial 4 μm-sections stained with hematoxylin and eosin (H&E) and submitted to immunohistochemical reactions. For the latter protocol, the sections were exposed to 366 antibodies against NeuN (neuronal marker; 1:1,000, clone A60, Merck Millipore, cat#MAB377, Temecula, CA, USA), MAP2 (neuronal marker; 1:1,000, clone M13, Thermo Fisher, cat#13-1500, Waltham, MA, USA), SMI 32 (neuronal marker; 1:2,500, clone SMI 32, 369 Biolegend, cat#SMI-32R, San Diego, CA, USA), GFAP (astrocytic marker, 1:100, clone 6F2, Dako/Agilent, cat#M0761, Santa Clara, CA, USA), vimentin (1:100, clone V9, e-Bioscience/Thermo Fisher, cat#14-9897-82, Waltham, MA, USA), CD34 (1:50, clone QBEnd-372 10, Dako, cat#M7165, Glostrup, Denmark) and CNPase (myeloarchitecture marker; 1:500, clone 11-5 B, Millipore, cat#MAB326, Darmstadt, Germany), for 18 h at 4°C. Then, a detection solution containing the secondary antibody and peroxidase (AdvanceTMHRP®, 375 Dako, cat#K4068, Glostrup, Denmark; or EnvisionTM Flex+, Dako, cat#K8002, Glostrup, Denmark) was added for 30 min at 37°C. 3,3-diaminobenzidine (DAB) was used as a chromogenic substrate and counterstaining was performed with hematoxylin. Negative 378 controls (without primary antibody) were run concurrently with all immunohistochemical reactions.

Samples were classified as FCD IIa or IIb according to the guidelines of the 381 International League Against Epilepsy (ILAE)[2]. Specifically, specimens exhibiting cortical dyslamination, hypertrophic and dysmorphic neurons [disoriented neurons with anomalous

cytoplasmic distribution of Nissl substance and accumulation of non-phosphorylated
384 neurofilament (SMI 32-positive)] without or with balloon cells (large cells with opaque eosinophilic cytoplasm, vesicular nucleus and immunopositivity for vimentin) were classified as FCD ILAE type IIa or IIb, respectively. Internal controls were derived from histologically
387 normal tissue where neither abnormal cells nor architectural changes were observed. Detailed patient information is provided in the Supplemental Table 1. All procedures were approved by the University of Campinas's Research Ethics Board, and written informed consent was
390 obtained from all patients before surgery.

Nuclei isolation and multiome library construction

393 The nuclei extraction protocol for human brain tissue was adapted from Jessa et al[44]. Briefly, frozen tissue was dounce homogenized and incubated for 2 minutes in lysis buffer containing 0.1% NP-40. Then, suspension was filtered through a 30 µm cell strainer (MACS), and washed
396 in buffer containing 10% BSA to remove debris, followed by a centrifugation at 500g and 4ºC for 5 minutes. Next, the suspension was carefully layered on double 29% and 50% iodixanol (Sigma) cushion layers, and centrifuged at 10,000g and 4ºC for 30 minutes for additional
399 removal of debris and cell aggregates. Retrieved nuclei were resuspended and counted. Nuclei were stained with Trypan blue and counted using a hemocytometer. Nuclei integrity was evaluated under a microscope using 40X or 60X magnifications, and nuclei were considered
402 viable if they were roundish and with an intact membrane. Next, we proceeded to the steps of the library construction, including open chromatin transposition and droplet (GEM) formation using a Chromium Controller equipment (10X Genomics). Libraries were prepared using the
405 Multiome ATAC + Gene Expression kit (10X Genomics) following the manufacturer's protocol CG000338 revision E. The quality of purified ATAC and RNA libraries was assessed using a TapeStation equipment (Agilent Technologies).

408

Sequencing and raw data processing

Multiome libraries were sequenced on either Illumina HiSeqX Ten or Illumina NovaSeq 6000
411 instruments according to the sequencing depth and the read length recommended in the Multiome kit protocol (CG000338). Demultiplexing, genome alignment, gene quantification, and peak accessibility in single nuclei were performed with the Cell Ranger ARC pipeline[45].
414 Specifically, fastq files were generated using cellranger mkfastq, then processed using the cellranger-arc count v.2.0.2 and the human genome GRCh38 to generate the count matrices for the RNA and ATAC modalities. Next, we performed data processing, normalization, and batch
417 integration on both assays independently, using Seurat v.5.0.1[46] and Signac v.1.12.0[47] in the R environment v.4.1.2.

420 **scRNA-seq analysis**

Samples were preprocessed in Seurat to remove low-quality nuclei using the following criteria: 400 < nFeatures < 7,000, 500 < UMI counts < 50,000, and percentage of mitochondrial reads
423 (percent.mt) < 15. Identification of doublets was performed per sample using scDblFinder v.1.8.0[48]. Next, standard Seurat data processing and normalization steps were performed: NormalizeData, FindVariableFeatures, ScaleData, RunPCA, RunUMAP, FindNeighbors, and
426 FindClusters. For each sample, the filtered counts matrix was log-normalized with regression of sequencing mitochondrial gene percentages. The RunUMAP function was used with the first 40 PCs identified in the elbow plot to perform dimensional reduction. Clustering was
429 performed using FindNeighbors and FindClusters functions using the smart local moving (SLM) algorithm, with a resolution of 1.2.

To correct batch effect among samples, we used Harmony v.1.2[24] to integrate sample-level
432 PCA projections using automatic hyperparameter optimization and default parameters.

Harmony implements an algorithm that projects cells into a common space where they cluster by cell type rather than other sample attributes such as sequencing batch.

435 Finally, we used SCTransform to create a normalized expression matrix containing all samples, and with regression of sequencing batch, brain region, and mitochondrial gene percentages. DEGs between FCD and controls were calculated using MAST or the Wilcox test implemented

438 in Seurat as indicated. For MAST, DEGs were selected using an adjusted P value of less than 0.05. The Wilcox test was applied using the SCT normalized data, using a log2 fold-change threshold of 0.25 and an adjusted P value of less than 0.05.

441

scATAC-seq analysis

Samples were preprocessed in Signac to remove low-quality nuclei using the following criteria:

444 1,000 < ATAC fragments < α(ATAC fragments), 400 < ATAC fragments in peaks < α (ATAC fragments in peaks), percent.mt < 15, nucleosome signal < 2, TSS enrichment < 2. Upper limits (α) of ATAC fragments per sample were computed as mean plus 2 standard deviations. Peaks

447 within each sample were identified using MACS2[49]. Standard processing and normalization steps were performed as follows: RunTFIDF, FindTopFeatures, RunSVD, and FindClusters. For each sample, the peak counts matrix was normalized using the RunTFIDF function, to

450 correct for differences in nuclei sequencing depth. Dimensional reduction was performed with latent semantic indexing (LSI) using the RunSVD function. UMAP projection was performed utilizing LSI components 2-30, as the first LSI component was confirmed to be a technical

453 variation. Before sample integration, we created a common peak set across samples using MACS2 and quantified peaks in each sample using the FeatureMatrix function. Peaks in non-standard chromosomes and blacklisted regions were removed from downstream analyses. We

456 then re-computed RunTFIDF, FindTopFeatures, and RunSVD on the merged object and applied Harmony to integrate sample-level low-dimensional cell embeddings (LSI components

2-30) using the same parameters defined in the RNA modality. Differentially accessible
459 chromatin regions (DACRs) between FCD tissues and controls were calculated using FindMarkers with the LR test, with a min.pct of 0.05, an adjusted P value less than 0.05, and included the number of ATAC fragments as a latent variable.

462

Multimodal data integration

To integrate gene expression data with chromatin accessibility data, we performed Weighted
465 Nearest Neighbor (WNN) analysis[25], an unsupervised approach that learns the relative utility of each modality and builds a joint graph representing a weighted combination of RNA and ATAC modalities. WNN was performed using the FindMultiModalNeighbors function with
468 20 neighbors (k.nn), based on the UMAP reductions obtained after Harmony integration. The WNN graph was used for creating a joint UMAP visualization and clustering based on the WNN graph was performed with the FindClusters function using the smart local moving
471 (SLM) algorithm. WNN clusters with fewer than 50 nuclei, a high percentage of doublets, or enriched in mitochondrial markers were removed from downstream analyses.

474 **Cluster annotation**

We used the Azimuth R package[25] to perform cell type annotation in the FCD dataset, using as a reference the Allen atlas of the human cortex[50]. Azimuth performed label-transferring by
477 assigning to each cell in the FCD dataset the most similar cell type in the reference based on gene expression similarity. The assigned cell type chosen for each cluster was defined as the cell type with the highest mean frequency within the cluster. The Azimuth annotation was
480 manually verified using the marker genes shown in Figure 1C. Cluster and subcluster markers were calculated using the function FindAllMarkers based on the SCT normalized data, with

the Wilcox test using a log2 fold-change threshold of 0.25 and an adjusted P value of less than
483 0.05.

Differential cellular abundance

486 We used the function propeller of the Speckle R package[26] to detect changes in cellular composition between FCD tissues and controls. The propeller function calculated cell type proportions in biological replicates, and fitted a linear model for each cell type with the tissue
489 condition as a covariate.

Integration with scRNA-seq from normal cortex

492 We obtained normal cortex scRNA-seq data from three autopsy donors with no history of neurological disease from Silleti et al[17]. To reduce computational costs, we randomly sampled 10,000 neuronal and 10,000 non-neuronal cortical nuclei from these donors, and processed this
495 data along with our FCD dataset following the procedures described in the "scRNA-seq analysis" section. We used Harmony v.1.2 to integrate all samples with automatic hyperparameter optimization and default parameters. Clustering was performed with
498 FindNeighbors and FindClusters functions using the smart local moving (SLM) algorithm and a resolution of 2.0.

501 **Gene signature analysis**

We used the AddModuleScore function from Seurat and the AUCell[32] R package to calculate gene set enrichment at the nuclei level. The AddModuleScore function from Seurat was
504 calculated based on the SCT-normalized RNA data. AUCell computes the gene set enrichment scores based on ranked gene lists (highest to lowest gene expression) in each cell. AUCell

scores were computed using the calcAUC function using the top 5% of genes. The obtained
507 threshold was used to classify cells with active status for the given signature.

Motif enrichment analysis

510 We performed motif analysis using chromVAR v. 3.3.2[31] and Signac. Motif activity at the nuclei level was calculated using the RunChromVAR function in Signac after matching the set of background peaks. Differential motif activity was computed using FindAllMarkers and
513 FindMarkers (pairwise comparisons) using the LR test, with a min.pct of 0.05, an adjusted P value less than 0.05, and included the number of ATAC fragments as a latent variable. We also filtered context-relevant motifs using the regulator expression levels. The regulator associated
516 with the motif needed to be expressed in at least 10% of the cells in the cluster.

Trajectory Inference Analysis

519 Pseudotime trajectory analysis was conducted using Monocle v3[51,52]. Briefly, RNA counts were preprocessed using the preprocess_cds function with the PCA method. To correct batch effects, we used the align_cds function[53] with the sample as an alignment group and a residual
522 formula including the sequencing batch, brain region, and mitochondrial genes percentage as covariates. Nuclei were then projected onto the UMAP space[54] and divided into clusters and partitions with the cluster_cells function[55]. Subsequently, an unsupervised trajectory was
525 identified using the SimplePPT (Probabilistic Pseudotime Trajectory) algorithm implemented in the learn_graph function[56], with the root set according to the expression of immature cellular markers. Pseudotime values were automatically assigned to individual cells based on their
528 distance along the trajectory relative to the root node. Genes with differential expression along the trajectory were calculated using the graph_test function, with adjusted P-value less than 0.05 and arranged according to Moran's I statistic. For each of these genes, differential

531 expression between tissue conditions was assessed using the fit_models and coefficient_table functions, using a regression test with an adjusted P value less than 0.05, minimum fold-change of 1.5, and with the number of counts, sequencing batch, brain region and mitochondrial gene
534 percentages as covariates.

Inference of regulatory links

537 We used a regression approach based on random forests to find open chromatin regions associated with gene regulation. The goal was to predict the expression levels of the target gene
540 based on the accessibility of ATAC peaks (i.e. cis-regulatory elements) around the gene promoter. Considering that a target gene *g* has *n* chromatin peaks, the linear regression was defined as:

543

$$x_{gi} = \sum_{i=1}^{n} w_j a_{ji} + \epsilon_i$$

546 where x_{gi} is the expression of *g* in cell *i*, a_{ji} is the accessibility of a given peak *j* in cell *i*, w_j is the weight and ϵ_i is the noise term. For the regression, we tested all peaks located in a 500kb window around the gene's transcriptional start site (TSS), to include both proximal and distal
549 regulatory elements. The optimal weights, which were associated with the importance of the cis-regulatory regions as predictors of gene expression, were determined using random forests. Specifically, the regression was implemented using the GENIE3 R package[32], setting the
552 number of trees to 200, and peaks with weights in the top 5% (95% percentile) for each gene were selected. We used the Encode annotation[57] downloaded from https: //screen.encodeproject.org to identify open chromatin regions with potential enhancer activity.
555 Encode enhancers are classified based on their relative location to the TSS as proximal

enhancer-like sequences (pELS) if located within 200bp < TSS ≤ 2kb, or distal enhancer-like sequence (dELS) when located more than 2kb from a TSS. We used both pELS and dELS datasets for annotation.

Acknowledgments

We thank the patients and their families who participated in this study. We thank Claudia Kleimann, Steven Hebert, Bhavyaa Chandarana and Tomas V. Waichman for advice on single-cell analysis; Damien Faury for sharing the nuclei extraction protocol; Danielle C. Bruno for discussions regarding the tissue dissociation. This work was supported by the following grants: Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP) Young Investigator grant 2019/07382-2 (D.F.T.V); the Chan Zuckerberg Initiative DAF, an advised fund of the Silicon Valley Community Foundation grant DAF2021-237598 (D.F.T.V.); and a FAPESP doctoral fellowship 2022/01530-2 (I.C.G). Authors also acknowledge the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) master fellowship (L.A.M.); Conselho Nacional de Pesquisa (CNPq) grant 311923/2019-4 (I.L.-C.); FAPESP doctoral fellowship 2020/06168-4 (J.C.G.), FAPESP regular grant (F.R.), and FAPESP Research Innovation and Dissemination Center grant 2013/07559-3 (F.C.).

Author contributions

I.C.G. and D.F.T.V. designed the study. L.A.M. and M.L. performed single-cell assays. I.C.G. and D.F.T.V. performed data analyses. P. A., J.C.G., C.L.Y., E.G., H.T., M.K.M.A., J.C.G, F.C., and I.L-C contributed to clinical sample collection. F.R. performed histopathological analyses. I.C.G. and D.F.T.V. wrote the manuscript. All authors reviewed and approved the final version of the manuscript.

Competing interests

The authors declare no competing interests.

585 **REFERENCES**

1. Blümcke, I. *et al.* The clinicopathologic spectrum of focal cortical dysplasias: A consensus classification proposed by an ad hoc Task Force of the ILAE Diagnostic Methods
588 Commission. *Epilepsia* **52**, 158–174 (2011).
2. Najm, I. *et al.* The ILAE consensus classification of focal cortical dysplasia: An update proposed by an ad hoc task force of the ILAE diagnostic methods commission. *Epilepsia*
591 **63**, 1899–1919 (2022).
3. Ii, P. H. I. & Crino, P. B. Focal Cortical Dysplasia: Gene Mutations, Cell Signaling, and Therapeutic Implications. *Annual Review of Pathology: Mechanisms of Disease* **12**, 547–
594 571 (2017).
4. Orlova, K. A. *et al.* Early Progenitor Cell Marker Expression Distinguishes Type II From Type I Focal Cortical Dysplasias. *J Neuropathol Exp Neurol* **69**, 850–863 (2010).
597 5. Hanai, S. *et al.* Abnormal maturation of non-dysmorphic neurons in focal cortical dysplasia: Immunohistochemical considerations. *Seizure* **19**, 274–279 (2010).
6. Lasarge, C. L. & Danzer, S. C. Mechanisms regulating neuronal excitability and seizure
600 development following mTOR pathway hyperactivation. *Frontiers in Molecular Neuroscience* **7**, (2014).
7. Powell, J. D., Pollizzi, K. N., Heikamp, E. B. & Horton, M. R. Regulation of Immune
603 Responses by mTOR. *Annual Review of Immunology* **30**, 39–68 (2012).
8. Weichhart, T., Hengstschläger, M. & Linke, M. Regulation of innate immune cell function by mTOR. *Nat Rev Immunol* **15**, 599–614 (2015).
606 9. Lin, Y. *et al.* Similar PDK1–AKT–mTOR pathway activation in balloon cells and dysmorphic neurons of type II focal cortical dysplasia with refractory epilepsy. *Epilepsy Research* **112**, 137–149 (2015).

609 10. Lim, J. S. *et al.* Brain somatic mutations in MTOR cause focal cortical dysplasia type II leading to intractable epilepsy. *Nat Med* **21**, 395–400 (2015).

11. Gerasimenko, A., Baldassari, S. & Baulac, S. mTOR pathway: Insights into an established
612 pathway for brain mosaicism in epilepsy. *Neurobiology of Disease* **182**, 106144 (2023).

12. Baulac, S. *et al. Single-Cell Genotyping and Transcriptomic Profiling in Focal Cortical Dysplasia*. https://www.researchsquare.com/article/rs-4014535/v1 (2024)
615 doi:10.21203/rs.3.rs-4014535/v1.

13. Boer, K. *et al.* Evidence of activated microglia in focal cortical dysplasia. *Journal of Neuroimmunology* **173**, 188–195 (2006).

618 14. Zimmer, T. S. *et al.* Balloon cells promote immune system activation in focal cortical dysplasia type 2b. *Neuropathol Appl Neurobiol* **47**, 826–839 (2021).

15. Galvão, I. C. *et al.* Identifying cellular markers of focal cortical dysplasia type II with cell-
621 type deconvolution and single-cell signatures. *Sci Rep* **13**, 13321 (2023).

16. Bakken, T. E. *et al.* Comparative cellular analysis of motor cortex in human, marmoset and mouse. *Nature* **598**, 111–119 (2021).

624 17. Siletti, K. *et al.* Transcriptomic diversity of cell types across the adult human brain. *Science* **382**, eadd7046 (2023).

18. Walter, T. J., Suter, R. K. & Ayad, N. G. An overview of human single-cell RNA sequencing
627 studies in neurobiological disease. *Neurobiology of Disease* **184**, 106201 (2023).

19. Liddelow, S. A. *et al.* Neurotoxic reactive astrocytes are induced by activated microglia. *Nature* **541**, 481–487 (2017).

630 20. Keren-Shaul, H. *et al.* A Unique Microglia Type Associated with Restricting Development of Alzheimer's Disease. *Cell* **169**, 1276-1290.e17 (2017).

21. Olah, M. *et al.* Single cell RNA sequencing of human microglia uncovers a subset associated with Alzheimer's disease. *Nat Commun* **11**, 6129 (2020).
22. Flynn, E., Almonte-Loya, A. & Fragiadakis, G. K. Single-Cell Multiomics. *Annual Review of Biomedical Data Science* **6**, 313–337 (2023).
23. Armand, E. J., Li, J., Xie, F., Luo, C. & Mukamel, E. A. Single-Cell Sequencing of Brain Cell Transcriptomes and Epigenomes. *Neuron* vol. 109 11–26 (2021).
24. Korsunsky, I. *et al.* Fast, sensitive and accurate integration of single-cell data with Harmony. *Nat Methods* **16**, 1289–1296 (2019).
25. Hao, Y. *et al.* Integrated analysis of multimodal single-cell data. *Cell* **184**, 3573-3587.e29 (2021).
26. Phipson, B. *et al.* propeller: testing for differences in cell type proportions in single cell data. *Bioinformatics* **38**, 4720–4726 (2022).
27. Kolberg, L. *et al.* g:Profiler—interoperable web service for functional enrichment analysis and gene identifier mapping (2023 update). *Nucleic Acids Research* **51**, W207–W212 (2023).
28. Baulac, S. *et al.* Single-cell genotyping and transcriptomic profiling in focal cortical dysplasia. Preprint at https://doi.org/10.21203/rs.3.rs-4014535/v1 (2024).
29. Paolicelli, R. C. *et al.* Microglia states and nomenclature: A field at its crossroads. *Neuron* **110**, 3458–3483 (2022).
30. Hafemeister, C. & Satija, R. Normalization and variance stabilization of single-cell RNA-seq data using regularized negative binomial regression. *Genome Biology* **20**, 296 (2019).
31. Schep, A. N., Wu, B., Buenrostro, J. D. & Greenleaf, W. J. chromVAR: inferring transcription-factor-associated accessibility from single-cell epigenomic data. *Nature Methods* **14**, 975–978 (2017).

32. Aibar, S. *et al.* SCENIC: single-cell regulatory network inference and clustering. *Nat Methods* **14**, 1083–1086 (2017).
657

33. Zhu, K. *et al.* Multi-omic profiling of the developing human cerebral cortex at the single-cell level. *Science Advances* **9**, eadg3754 (2023).

34. Baysoy, A., Bai, Z., Satija, R. & Fan, R. The technological landscape and applications of single-cell multi-omics. *Nat Rev Mol Cell Biol* **24**, 695–713 (2023).
660

35. Schirmer, L. *et al.* Neuronal vulnerability and multilineage diversity in multiple sclerosis. *Nature* **573**, 75–82 (2019).
663

36. Pfisterer, U. *et al.* Identification of epilepsy-associated neuronal subtypes and gene expression underlying epileptogenesis. *Nat Commun* **11**, 5038 (2020).

37. Khalil, M. *et al.* Neurofilaments as biomarkers in neurological disorders — towards clinical application. *Nat Rev Neurol* **20**, 269–287 (2024).
666

38. Mathys, H. *et al.* Single-cell transcriptomic analysis of Alzheimer's disease. *Nature* **570**, 332–337 (2019).
669

39. Masuda, T. *et al.* Spatial and temporal heterogeneity of mouse and human microglia at single-cell resolution. *Nature* **566**, 388–392 (2019).

40. Potru, P. S. & Spittau, B. CD74: a prospective marker for reactive microglia? *Neural Regen Res* **18**, 2673–2674 (2023).
672

41. Kumar, P. *et al.* Single-cell transcriptomics and surface epitope detection in human brain epileptic lesions identifies pro-inflammatory signaling. *Nat Neurosci* **25**, 956–966 (2022).
675

42. Lattke, M. & Guillemot, F. Understanding astrocyte differentiation: Clinical relevance, technical challenges, and new opportunities in the omics era. *WIREs Mech Dis* **14**, e1557 (2022).
678

43. Verhoog, Q. P., Holtman, L., Aronica, E. & van Vliet, E. A. Astrocytes as Guardians of Neuronal Excitability: Mechanisms Underlying Epileptogenesis. *Front. Neurol.* **11**, (2020).

681 44. Jessa, S. *et al.* Stalled developmental programs at the root of pediatric brain tumors. *Nat Genet* **51**, 1702–1713 (2019).

45. Zheng, G. X. Y. *et al.* Massively parallel digital transcriptional profiling of single cells. *Nat*

684 *Commun* **8**, 14049 (2017).

46. Dictionary learning for integrative, multimodal and scalable single-cell analysis | Nature Biotechnology. https://www.nature.com/articles/s41587-023-01767-y.

687 47. Stuart, T., Srivastava, A., Madad, S., Lareau, C. A. & Satija, R. Single-cell chromatin state analysis with Signac. *Nat Methods* **18**, 1333–1341 (2021).

48. Germain, P.-L., Lun, A., Garcia Meixide, C., Macnair, W. & Robinson, M. D. Doublet

690 identification in single-cell sequencing data using scDblFinder. *F1000Res* **10**, 979 (2021).

49. Zhang, Y. *et al.* Model-based Analysis of ChIP-Seq (MACS). *Genome Biology* **9**, R137 (2008).

693 50. Comparative cellular analysis of motor cortex in human, marmoset and mouse | Nature. https://www.nature.com/articles/s41586-021-03465-8.

51. Trapnell, C. *et al.* The dynamics and regulators of cell fate decisions are revealed by

696 pseudotemporal ordering of single cells. *Nat Biotechnol* **32**, 381–386 (2014).

52. Cao, J. *et al.* The single-cell transcriptional landscape of mammalian organogenesis. *Nature* **566**, 496–502 (2019).

699 53. Haghverdi, L., Lun, A. T. L., Morgan, M. D. & Marioni, J. C. Batch effects in single-cell RNA-sequencing data are corrected by matching mutual nearest neighbors. *Nat Biotechnol* **36**, 421–427 (2018).

54. McInnes, L., Healy, J. & Melville, J. UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction. Preprint at http://arxiv.org/abs/1802.03426 (2020).

55. Traag, V. A., Waltman, L. & van Eck, N. J. From Louvain to Leiden: guaranteeing well-connected communities. *Sci Rep* **9**, 5233 (2019).

56. Qiu, X. *et al.* Reversed graph embedding resolves complex single-cell trajectories. *Nat Methods* **14**, 979–982 (2017).

57. Moore, J. E. *et al.* Expanded encyclopaedias of DNA elements in the human and mouse genomes. *Nature* **583**, 699–710 (2020).

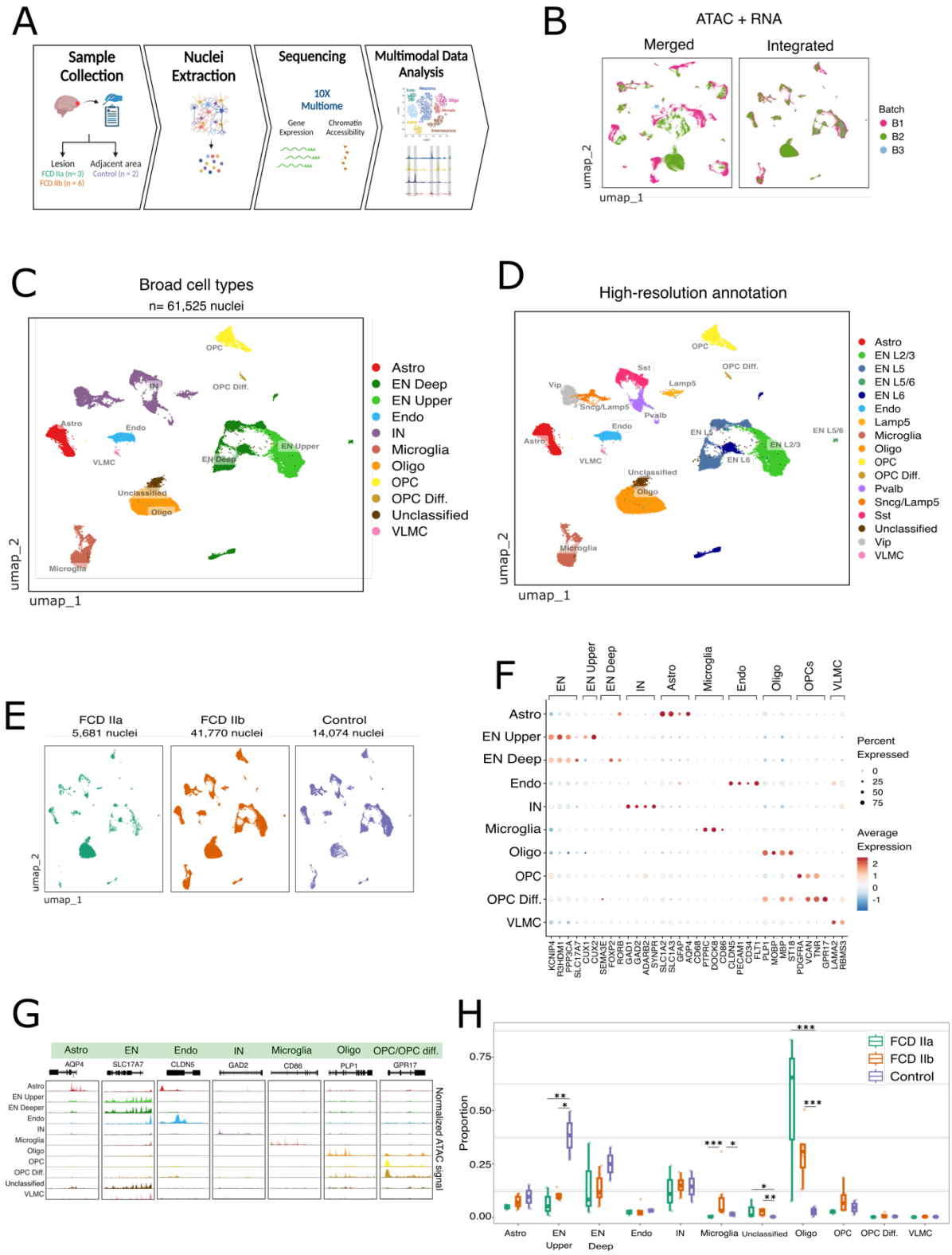


Figure 1. Overview of the multimodal single-cell sequencing in FCD type II.

(A) Schematic of FCD samples and study design. Cortical tissues from lesions and adjacent non-lesion areas in individuals with FCD IIa and IIb were profiled by multiome single-cell ATAC and RNA sequencing. Created with Biorender.com.

(B) Uniform manifold approximation and projection (UMAP) visualization of joint modalities
720 (ATAC + RNA) after sample integration. Nuclei are colored by sequencing batch.

(C) UMAP joint visualization of nuclei colored by broad cell type annotation.
723

(D) UMAP joint visualization colored using a high-resolution annotation with classification of neuronal subtypes.
726

(E) UMAP plots depicting nuclei by tissue condition.

729 **(F)** Dot plot showing gene expression of known marker genes for major cortical cell types. Dot size corresponds to the fraction of cells expressing the marker, and color corresponds to the expression levels.
732

(G) Coverage plots of marker genes depicting chromatin accessibility across cell types. The normalized ATAC signal is depicted in a region of +/- 1kb around gene coordinates.
735

(H) Box plots showing cell type proportion by tissue condition. In box plots, the median is indicated by the center line; box limits represent upper and lower quartiles; and whiskers extend
738 to 1.5 times the interquartile range. Overlay dots represent cell type proportions in individual samples. Significant changes in cell types were detected using Speckle. ˙P < 0.1, *P < 0.5, **P < 0.01, ***P < 0.001.
741

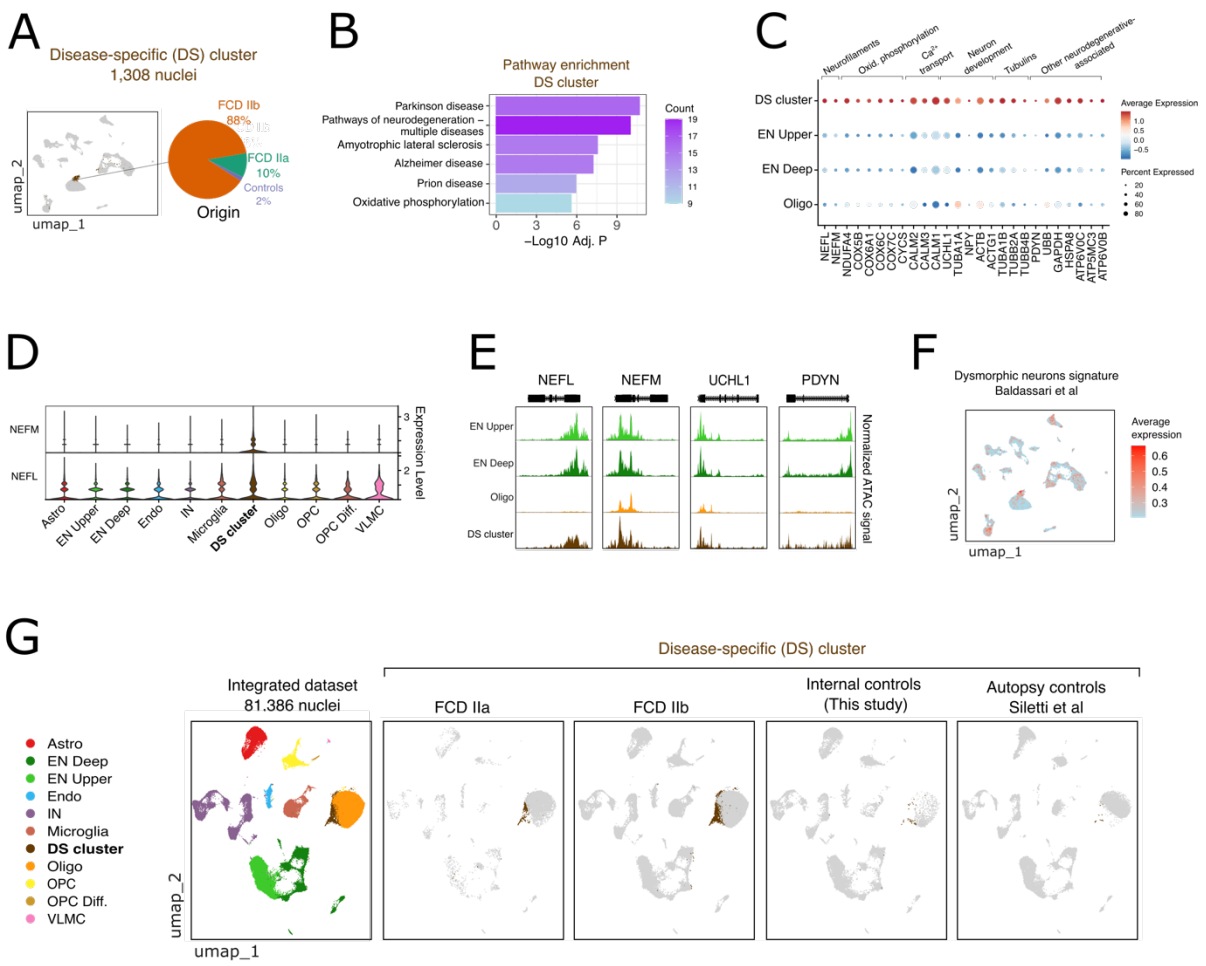


Figure 2. Characterization of a disease-specific cluster containing dysmorphic neurons.

(A) UMAP highlighting the disease-specific (DS) cluster. The pie chart indicates the percentage of nuclei from FCD IIa, IIb, and controls within the cluster.

(B) KEGG pathways enrichment of DS cluster markers. The bars represent the significance score of pathway enrichment, and the color indicates the number of genes in each pathway.

(C) Dot plot denoting expression of marker genes associated with neurons in the DS cluster. Neuronal genes are grouped by biological function. Dot size and color correspond to expression frequency and level in the cluster, respectively.

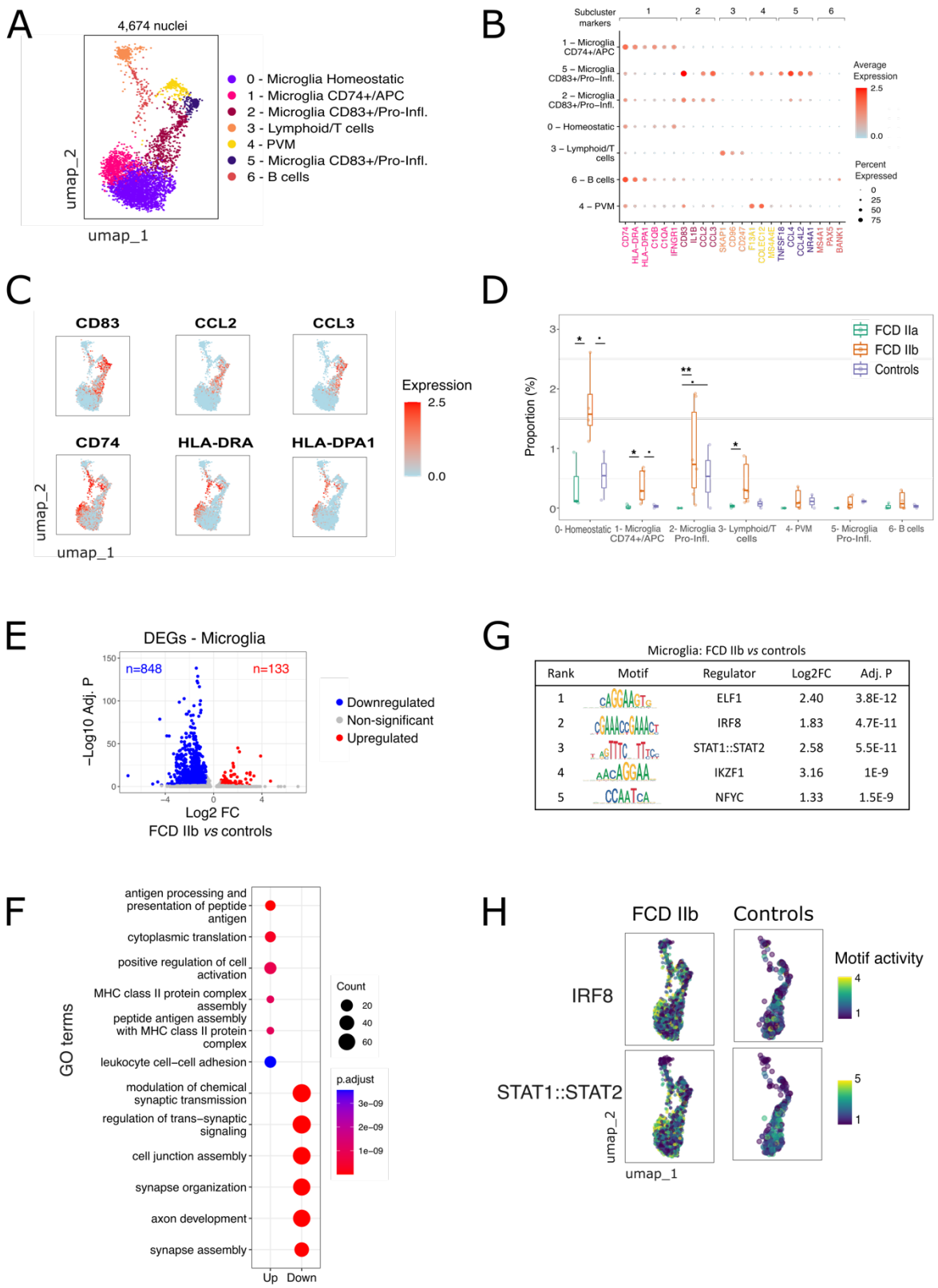
(D) Violin plots denoting gene expression of *NEFM* and *NEFL* neurofilaments in the various cell types.

(E) Coverage plots of neuronal genes expressed in the DS cluster. The normalized ATAC signal is depicted in a region of +/- 1kb around gene coordinates.

(F) UMAP showing the module score of a dysmorphic neuron (DN) signature obtained from Baldassari et al.

(G) UMAP RNA plots showing the DS cluster population in FCD IIa and IIb lesions, histologically normal tissue from FCD individuals (internal controls), and normal cortex
768 (autopsy controls). The leftmost UMAP depicts the cell types in the integrated dataset, which was created by integrating data from this study with Siletti et al.

771



774 **Figure 3 – Microglia pathological cell states and activation in FCD type IIb.**

(A) UMAP visualization of microglia and lymphoid subclusters sequenced in FCD lesions.

777

(B) Feature plots depicting expression of key genes distinguishing microglia subsets *CD83*$^{+}$ (*CD83*, *CCL2*, *CCL3*) and *CD74*$^{+}$ (*CD74*, *HLA-DRA*, *HLA-DPA1*).

780

(C) Dot plot denoting the expression of key marker genes in microglia and lymphoid subclusters.

783

(D) Box plots denoting subcluster frequency by tissue condition. Significant changes were detected using Speckle. Box plots and significance legends are defined in Fig. 1H.

786

(E) Volcano plot depicting differentially expressed genes in FCD IIb microglia. The *x*-axis is the fold-change (log2) expression in FCD IIb *vs* controls, and the *y*-axis depicts the significance

789 of the change (-log10 of the adjusted P-value). The color indicates the gene status.

(F) Dot plots showing top enriched GO terms in up- and down-regulated genes from (E). Dot

792 size and color correspond to the number of genes associated with the GO term and enrichment P-value, respectively.

795 **(G)** Top active motifs in FCD IIb microglia open chromatin regions, as measured by chromVAR.

798 **(H)** Feature plots depicting IRF8 and STAT1/STAT2 motif activity in the microglia from FCD IIb and controls.

801

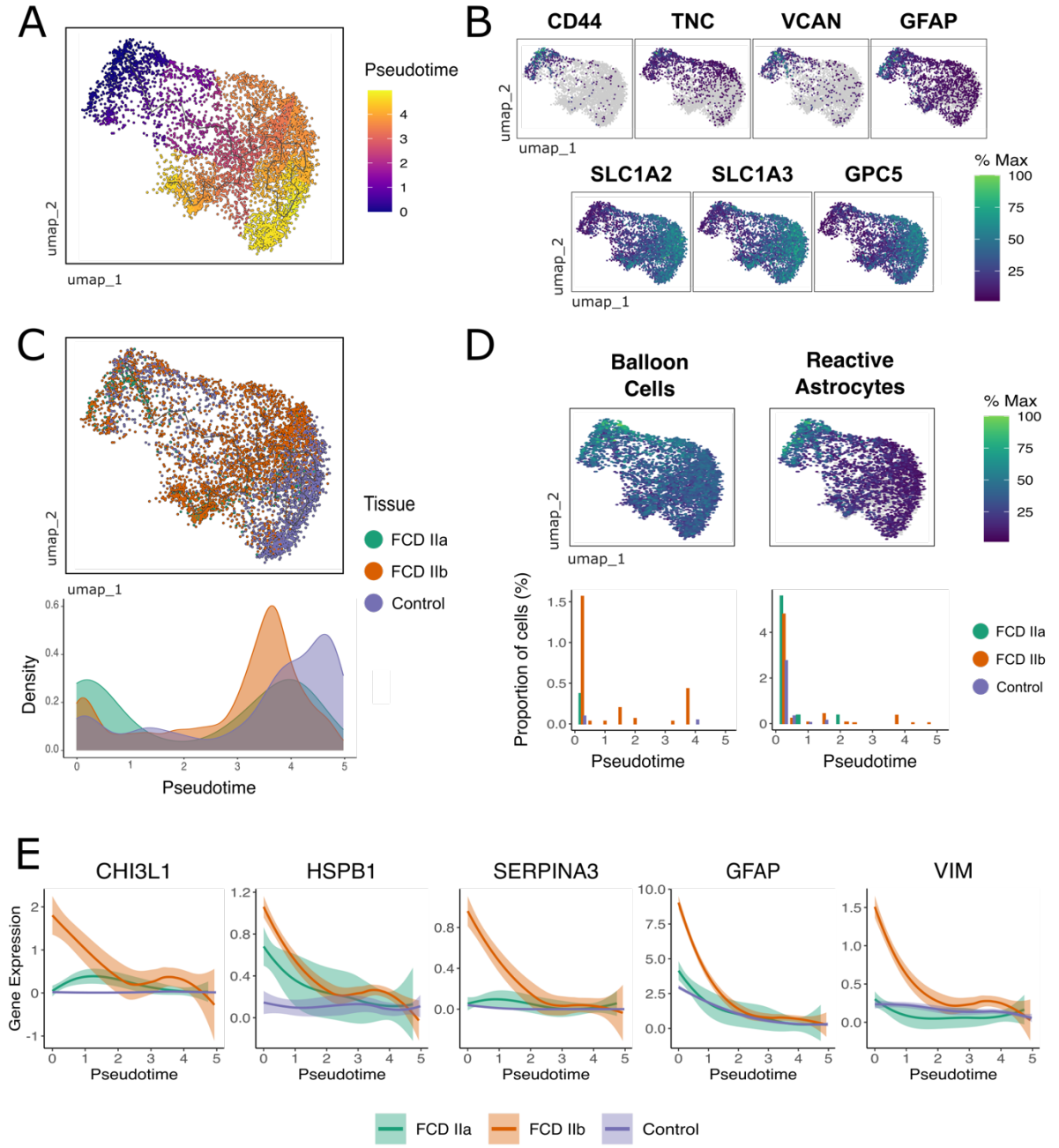


Figure 4. Impaired Astrocyte Differentiation in FCD.

(A) UMAP visualization of astrocytes according to pseudotime trajectory estimated by Monocle3.

(B) Feature plots depicting gene expression of immature astrocyte markers *CD44*, *TNC*, *VCAN*, and *GFAP* (top), and differentiated astrocyte markers *SLC1A2*, *SLC1A3*, and *GPC5* (bottom).

(C) UMA visualization and density plot quantification of tissue condition along the trajectory.

(D) Feature plots depicting expression of balloon cells and reactive astrocytes gene signatures. Bar graphs (at the bottom) are quantifying active cells for these signatures along the pseudotime
816 trajectory, and cells are colored by tissue condition.

(E) Expression levels of genes associated with balloon cells and reactive astrocytes *CHI3L1*,
819 *HSPB1*, *SERPINA3*, *GFAP*, and *VIM* in the trajectory and by tissue condition.

822

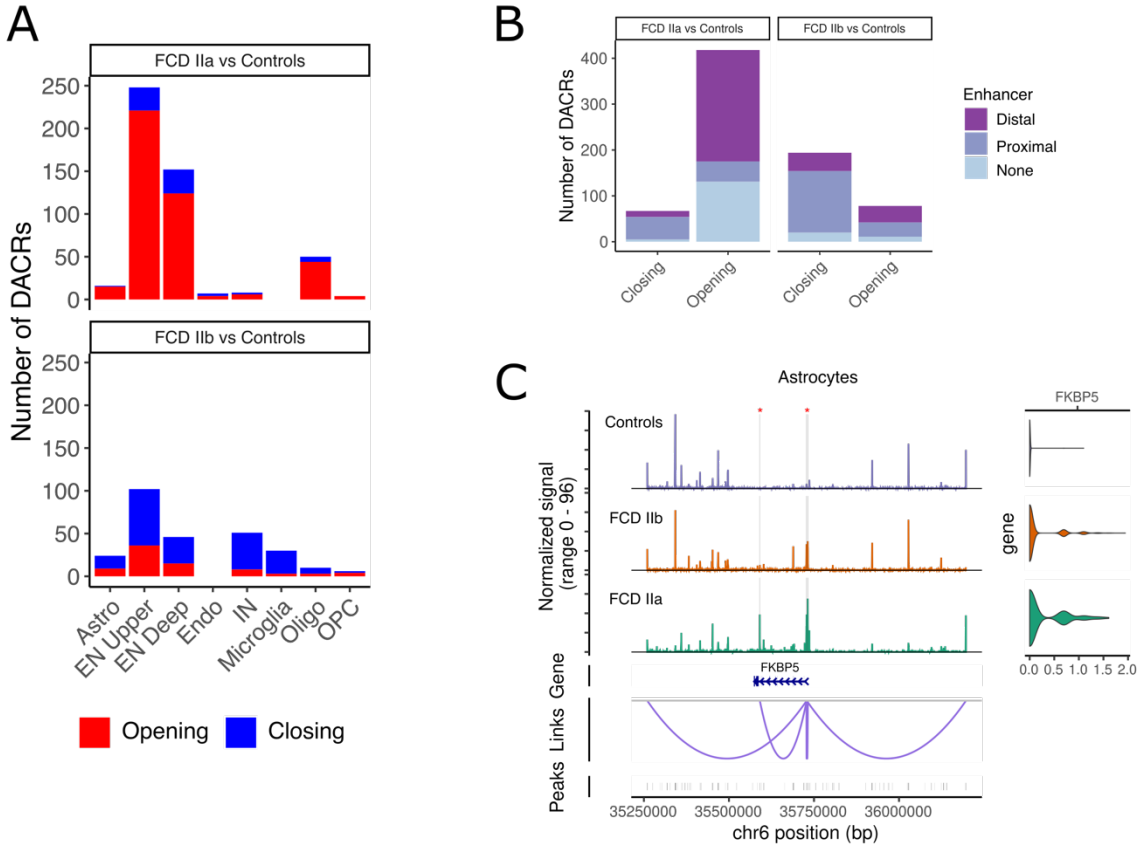
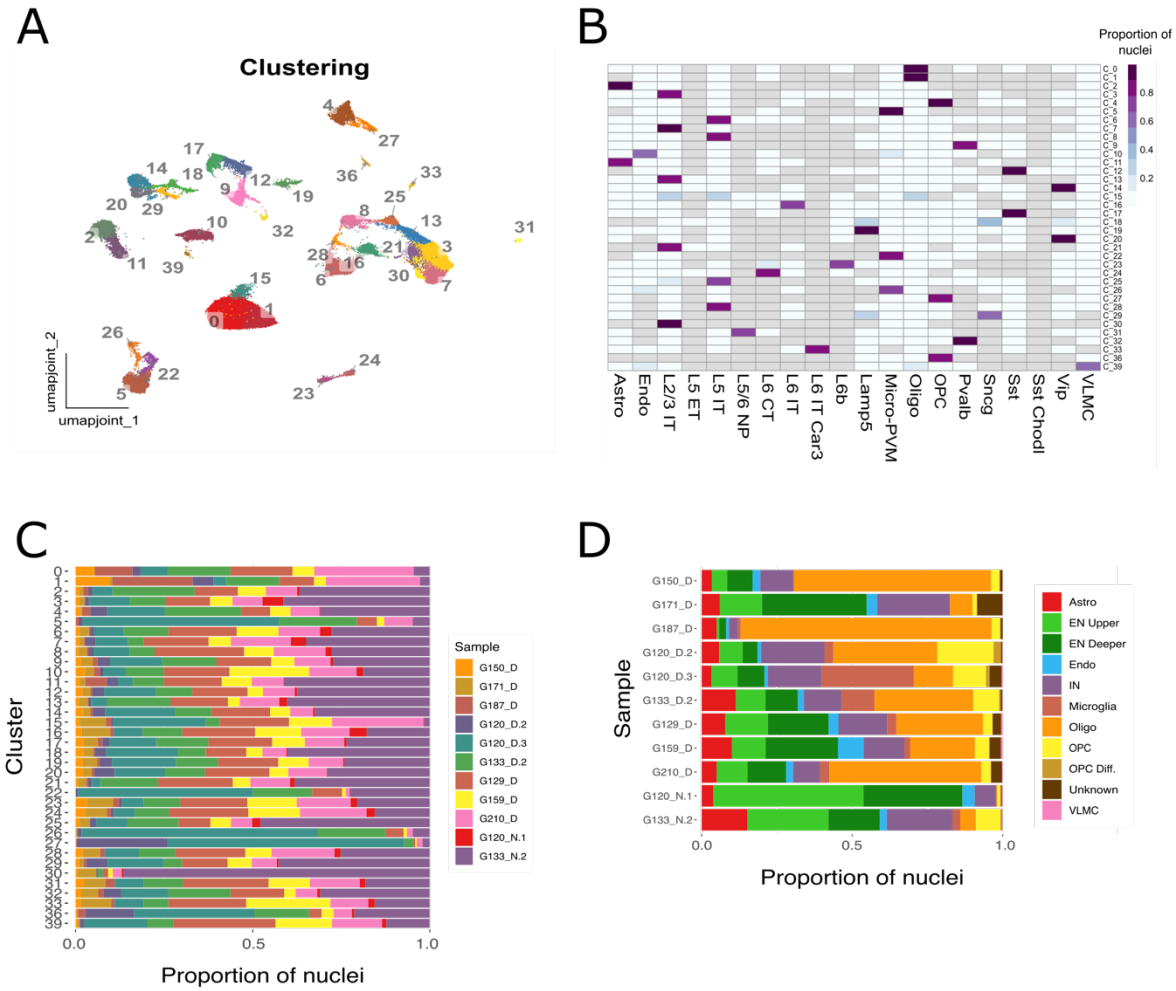


Figure 5. Open chromatin changes in FCD and *FKBP5* regulation in astrocytes.

(A) Barplots depicting differentially accessible chromatin regions (DACRs) between FCD tissues and controls. The number of chromatin regions with opening or closing status (gain or loss of chromatin accessibility, respectively) is indicated for each cell type.

(B) Enhancer types overlapping DACRs. The number of opening and closing peaks associated with distal or proximal enhancers from Encode are indicated.

(C) Regulatory links predicted for *FKBP5* in astrocytes. ATAC coverage tracks and corresponding expression levels in controls, FCD IIb, and IIa tissues are depicted at the top. Regions highlighted in gray indicate opening peaks in lesions and red asterisks indicate Encode-annotated enhancers. The gene track shows the location of the *FKBP5* locus in chromosome 6, and the bottom track shows all peaks contained in the chromosomal region.



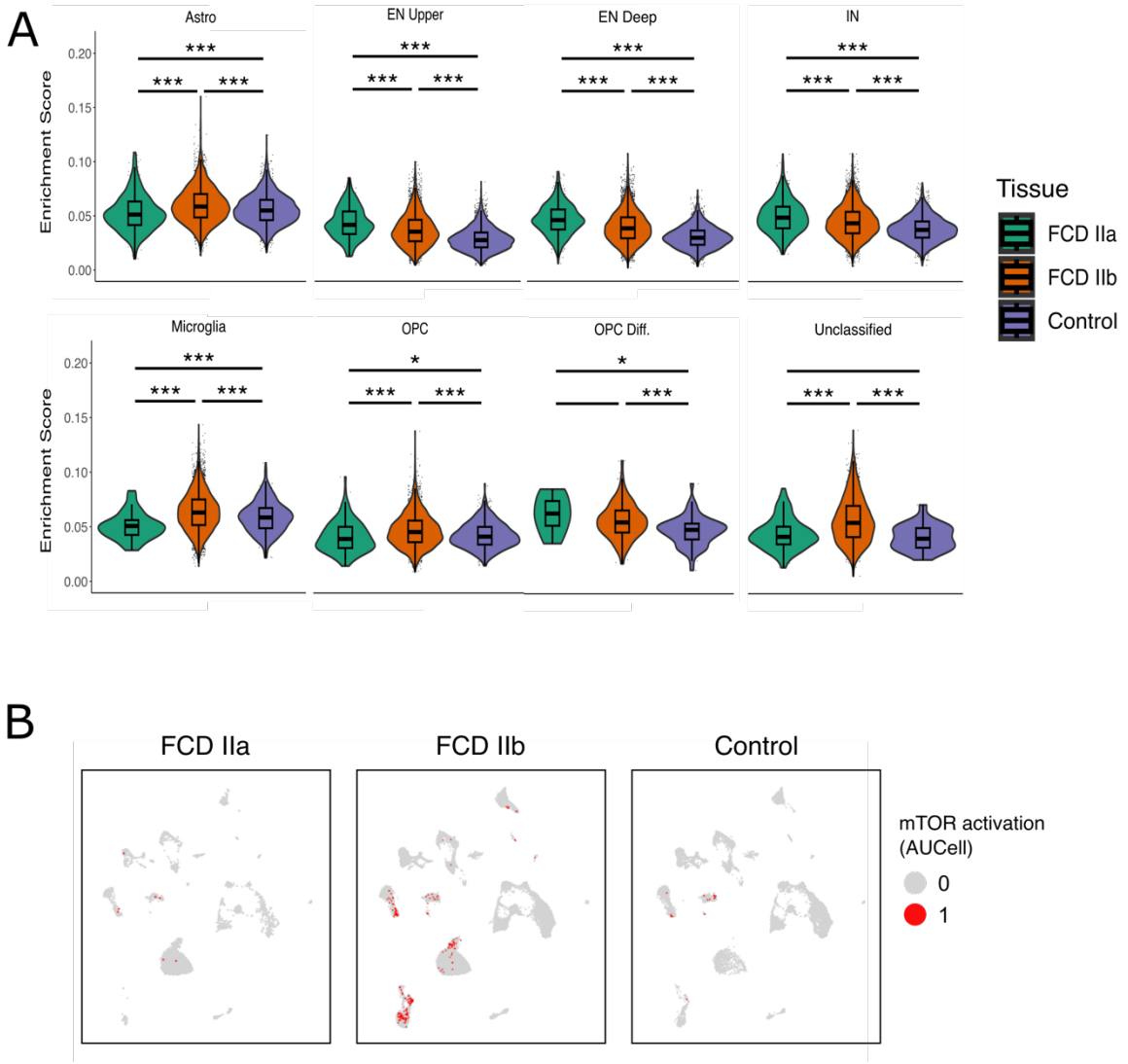
Supplementary Figure 1. Clustering and annotation.

(A) UMAP visualization of nuclei clusters in the multimodal FCD dataset. Clustering was performed in the joint graph (ATAC + RNA) obtained by Weighted Nearest Neighbor (WNN) analysis.

(B) Heatmap of cluster annotation performed by Azimuth based on the Allen atlas of the human cortex. The color indicates the proportion of Azimuth subtypes identified in each cluster.

(C) Bar graphs denoting the distribution of samples in each cluster.

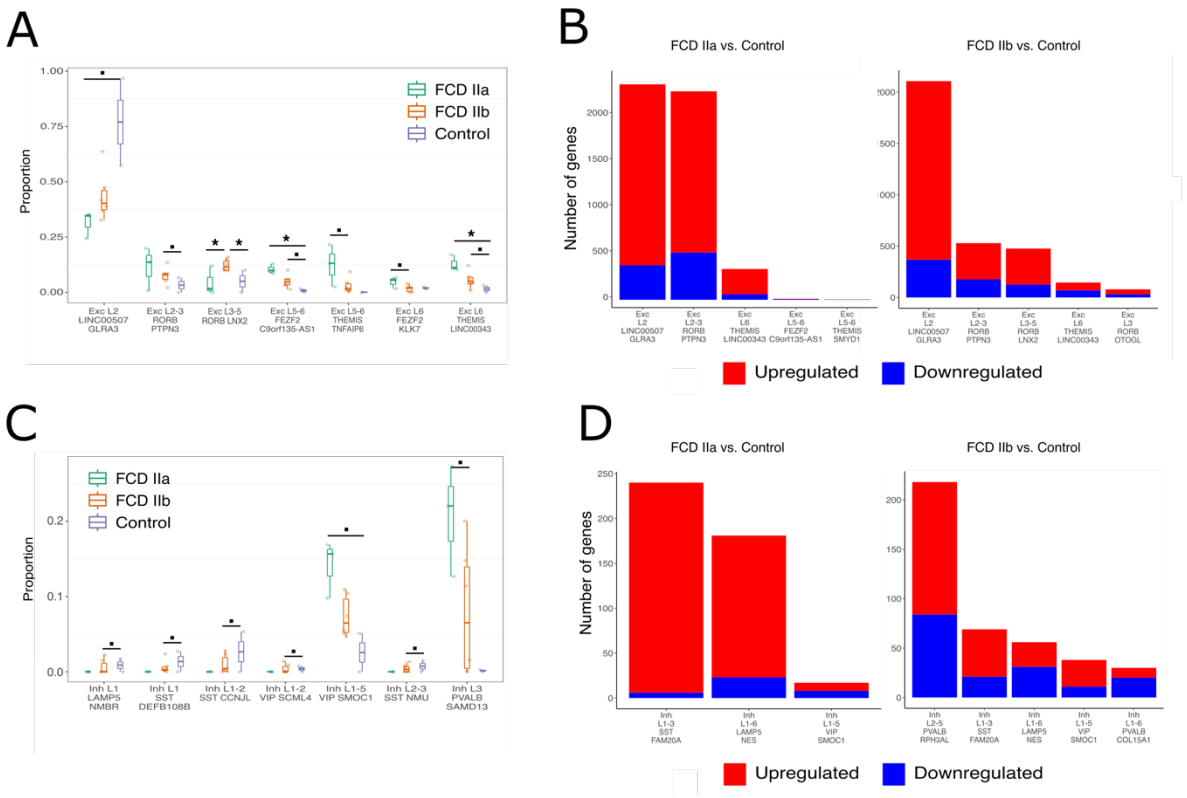
(C) Bar graphs showing the distribution of final consensus cell types across nuclei clusters.



Supplementary Figure 2. Expression of mTORC1 signaling across cell types in FCD lesions.

(A) Violin plots of the mTORC1 signaling enrichment score by cell type and tissue condition. The enrichment score was calculated using the AddModuleScore function from Seurat, using the mTORC1 signature from MSigDB (https://www.gsea-msigdb.org/gsea/msigdb). In box plots, the median is indicated by the center line; box limits represent upper and lower quartiles; and whiskers extend to 1.5 times the interquartile range. Significant changes in cell types were detected using a Wilcox ran-sum test. *P < 0.5, **P < 0.01, ***P < 0.001.

(B) Feature plots showing cells identified as active for the mTORC1 signature by the AUCell method. Cells with scores higher and lower than the AUCell threshold are shown in red and gray, respectively.



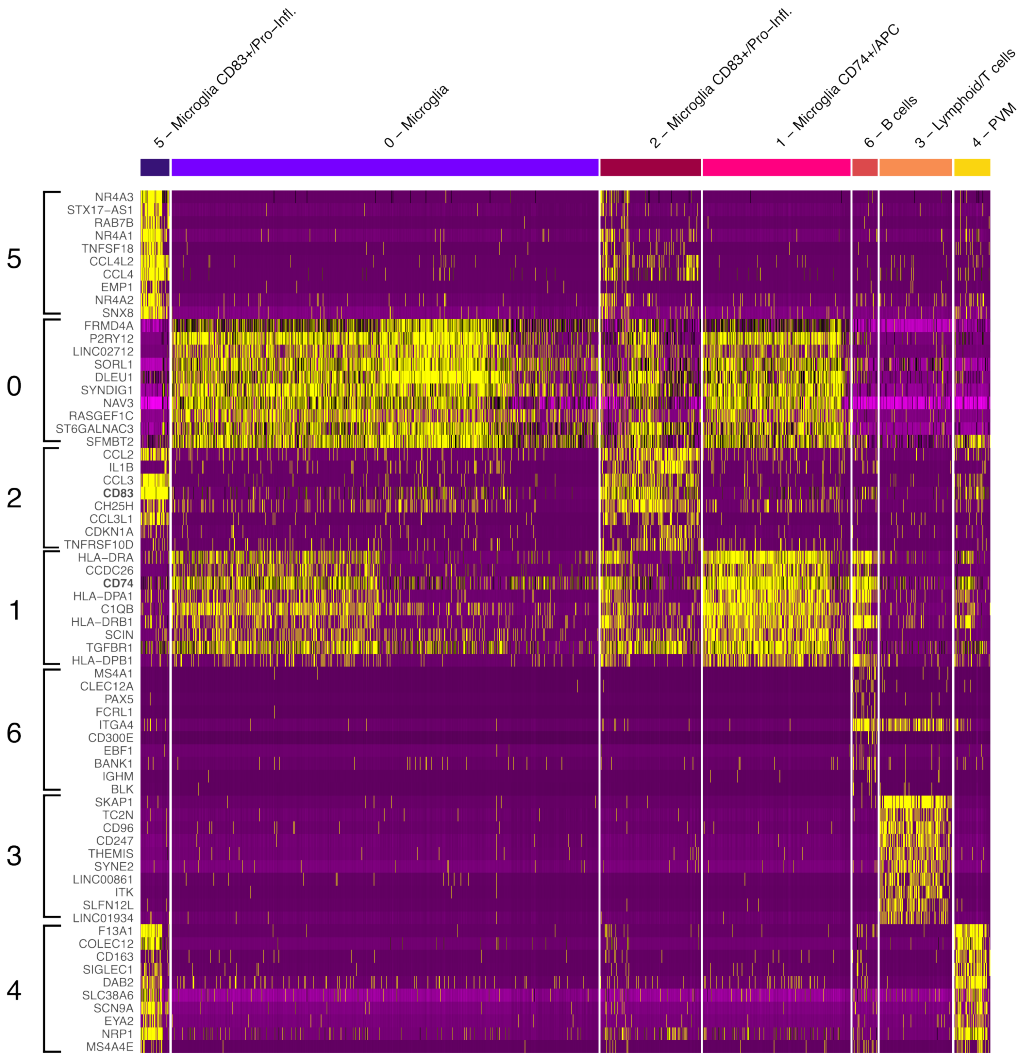
Supplementary Figure 3. Excitatory and inhibitory neuron subtypes affected in FCD.

(A) Box plots showing excitatory neurons (EN) subtype frequency in FCD and controls. In box plots, the median is indicated by the center line; box limits represent upper and lower quartiles; and whiskers extend to 1.5 times the interquartile range. Overlay dots represent proportions in individual samples. Significant changes were detected using Speckle. ˙P < 0.1, *P < 0.5, **P < 0.01, ***P < 0.001.

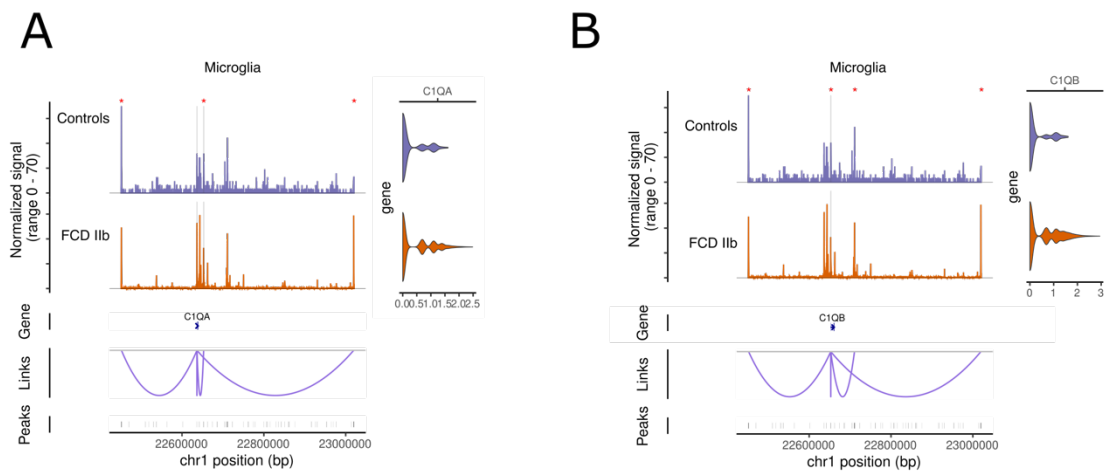
(B) Bar graphs depicting the number of differentially expressed genes (DEGs) in EN subtypes. DEGs between FCD and controls were computed using MAST with adj. P < 0.05. Only the five subtypes with the most transcriptional changes are shown.

(C) Box plots showing inhibitory neurons (IN) subtype frequency in FCD and controls. Box plots legends and significance levels are defined as in (A).

(D) Bar graphs depicting DEGs in IN subtypes. DEGs between FCD and controls were computed using MAST with adj. P < 0.05. Only the five subtypes with the most transcriptional changes are shown.



Supplementary Figure 4. Heatmap depicting expression of top 10 markers in microglia subclusters. Markers were computed using the function FindMarkers from Seurat using the Wilcox test, with a min.pct of 0.1, adjusted $P < 0.05$, and log2 fold-change > 1.5.



Supplementary Figure 5. Open chromatin regions linked to *C1QA* and *C1QB* regulation FCD IIb microglia.

(A) Regulatory links predicted for *C1QA* in microglia. ATAC coverage tracks and corresponding expression levels in controls and FCD IIb tissues are depicted at the top. Regions highlighted in gray indicate opening peaks in FCD and red asterisks indicate Encode-annotated enhancers. The gene track shows the location of the *C1QA* locus in chromosome 1, and the bottom track shows all peaks contained in the chromosomal region.

(B) Regulatory links predicted for *C1QB* in microglia. Legends are defined as in (A) above.